

Semantic Representation

Lenhart Schubert

University of Rochester
Rochester, NY 14627-0226

Abstract

In recent years, there has been renewed interest in the NLP community in genuine language understanding and dialogue. Thus the long-standing issue of how the semantic content of language should be represented is reentering the communal discussion. This paper provides a brief “opinionated survey” of broad-coverage semantic representation (SR). It suggests multiple desiderata for such representations, and then outlines more than a dozen approaches to SR—some long-standing, and some more recent, providing quick characterizations, pros, cons, and some comments on implementations.

Introduction

There are signs of a resurgence of interest in genuine language understanding within NLP and AI. These signs include various challenges that can be viewed as scaled-back versions of the Turing Test, in particular RTE (Recognizing Textual Entailment) (Dagan, Glicksman, and Magnini 2006), COPA (Choice of Plausible Alternatives) (Roemmele, Bejan, and Gordon 2011), WSC (Winograd Schema Challenge) (Levesque, Davis, and Morgenstern 2012), and the Aristo Challenge (Clark 2015), as well as a recent ACL workshop on Semantic Parsing (Artzi, Kwiatkowski, and Berant 2014), and applications such as natural language (NL) dialogue with robots (Eliasson 2007; Howard, Tellex, and Roy 2013) and semantics-guided MT (Jones et al. 2012). Thus the long-standing issue of how the semantic content of language should be represented is reentering the communal discussion. The focus here will be on broad-coverage semantic representation.

Desiderata

In a sense, for a representation to be meaningful to a computer, it need only have the right computational properties, given the tasks the computer is performing. But that is not enough. As builders of potentially very large, complex systems dependent on symbolic representations, we also have to be concerned with the naturalness of the representations from *our* perspective. It should be easy to see

intuitively whether a putative semantic interpretation of a given sentence (in context) really *does* capture its semantic content. The following desiderata seem to me like natural ones, falling into both categories (and all of them have been considered explicitly or tacitly by various researchers at various times).

- *Language-like expressivity of the semantic representation (SR)*: All languages allow for the devices familiar from first-order logic: predicates + connectives + quantifiers + equality. But they allow more, including: generalized quantifiers (*most men who smoke*), intensional predicates (*believe, intend, resemble*), predicate and sentence modification (*very, gracefully, nearly, possibly*), predicate and sentence reification (*Birds evolved from dinosaurs, Beauty is subjective, That exoplanets exist is now certain*), and reference to events or situations (*Many children had not been vaccinated against measles; this situation caused sporadic outbreaks of the disease*).

High expressivity is sometimes claimed to be undesirable, on the grounds that it is an obstacle to efficient inference. That is a reasonable stance for designers of specialized systems that simply don't *require* high expressivity. But it is unreasonable on two counts for a general semantic representation (which is our focus here): One is that efficient inference methods for expressively limited subsets of an expressively rich representation are easily embedded within the richer representation (as specialized reasoning subsystems); the other is that the richer representation will enable inferences that are not possible at all, or at best much more cumbersome, in a restricted representation. An analogous, equally unreasonable restriction in programming language design would be to prohibit recursion, runtime computation of iteration bounds, and backward branching, as a way of assuring efficient execution. Much power would be lost in doing so.

- *Ease of transduction between surface structure and SR*, irrespective of the domain (i.e., modular, easily understood, easily edited representations of the transductions); Richard Montague's work (Dowty 1979) revealed a beautiful, direct correspondence between phrase structure and meaning structure, one that accounts for entailments (or their absence) even in intensional contexts. It would be astounding if this correspondence, which is the basis of

compositional semantics and interpretation, were a mere cultural accident.

- *Availability of referents for anaphoric expressions*; pronouns and definite noun phrases can refer to a wide variety of entities that were previously introduced explicitly or implicitly (including propositions, events, situations, actions, types of actions, kinds of entities and actions, and more); an adequate SR should make these referents available.
- *Accord with semantic intuitions*; we have many quite compelling intuitions about the meanings of the words we use (else lexicography would not be possible), about alternative interpretations of given phrases or sentences, and about the entailments of sentences or smaller linguistic units. A SR should do justice to these intuitions.
- *Ease of use for inference*; humans appear to perform numerous inferences both during and as a result of the language understanding process, and can answer questions inferentially about the proffered input, in concert with background knowledge; SRs should enable similar inferring.
- (Related) *Ease of integration with specialized methods* for taxonomies, times, paronomies, spatial/imagistic representations, collections, symbolic entities, math, procedures, etc.; this is needed if we wish to match human proficiency in these pervasive domains.
- *Formal interpretability*; model theory formally specifies what kinds of entities in a domain of discourse are allowed as semantic values (denotations) of our SR vocabulary, and how we can derive the semantic values of more complex expressions from those of the basic vocabulary. As such it provides a tool to the SR designer for ensuring semantic coherence and consistent use of the SR; in addition, it provides a way of determining whether proposed inference methods will yield true conclusions from true premises, or at least yield plausible conclusions (depending on the kinds of inferences we are after).

That said, it should be noted that historically, the development of full formal semantics for Boolean logic, first-order logic, modal logics, and other more esoteric logics has lagged (often by decades) behind their syntactic development (including methods of proof), yet those syntactic developments rarely ran astray; careful intuitive consideration of the intended meaning of new constructs, guided by analogy with their linguistic counterparts, generally led to sensible results. Much the same can be said about the evolution of nonmonotonic and uncertain inference in AI. So we should be cognizant of the benefits of model theory, without depriving ourselves of essential expressive devices for lack of a model theory that would satisfy the editors of the *Notre Dame Journal of Formal Logic*. Indeed it is entirely possible that the “right” semantic representation for mechanizing human language understanding and use necessarily retains a certain NL-like looseness that thwarts full formalization.
- *Trainability, ease of semantic rule learning*; we wish to be able to improve the performance of systems that map be-

tween language and the SR using machine learning techniques, and indeed to learn the forms of the transduction rules; the latter may be easy or hard, depending on the directness of the correspondence between language and the SR – see the second bullet above.

Approaches

We now briefly characterize a wide variety of approaches to SR, pointing out pros and cons and occasionally mentioning extant implementations. Since the characterizations often group together multiple variants of a general approach, not everything said here applies to all members of a group, nor are all major variants cited. But hopefully the comments do capture enough of the distinctive “flavor” of each type of approach to render the cited pros and cons comprehensible.

Structured English; with polarity (and perhaps other) annotations; a very simple, direct “semantic representation”!

Pros: Traditionally, reliance on ordinary language has been considered antithetical to trustworthy reasoning; but a surprising recent development has been the successful implementation of *Natural Logic (NLog)* for obtaining obvious lexical entailments, such as, *Jimmy Dean refuses to move without his blue jeans* \models *James Dean won't dance without pants* (MacCartney and Manning 2009; Angeli and Manning 2014). The three main principles used in NLog are the following: (i) Certain complement-taking lexical items (e.g., *refuse*, *without*) flip the polarity of their complement; (ii) we can replace a lexical item by a more general (more specific) one in a positive-polarity (negative-polarity) one; and (iii) certain *implicative* verbs (e.g., *manage*, *refuse*) entail or at least strongly suggest the truth (or falsity) of their complement. In the example, *Jimmy Dean* can be replaced by the synonym *James Dean*; *move* can be replaced by the more specific term *dance* in the negative environment formed by *refuse*; *blue jeans* can be replaced by the more general term *pants* in the doubly negative, hence positive environment created by *refuse* and *without*; and *refuse* entails the negation of its complement, i.e., from *refuse to V* we can infer *won't V*.

A related approach to inference has come out of RTE research, making use of (soft) entailment relations between syntactic subtrees (Dagan et al. 2008; Berant, Dagan, and Goldberger 2011; Melamud et al. 2013). In fact a new area of statistical NL semantics has sprung up based on the idea that words are semantically similar if they tend to co-occur with the same neighboring words. Thus statistical semantics endeavors to reconstrue logical algebra as vector algebra, where words are high-dimensional vectors whose components are frequencies of collocation with other words or groups of words, and longer constituents are represented as vector compositions (Widdows 2004; Clarke 2012; Clark 2013).

If nothing else, the developments in the use of NLog and in statistical compositional semantics for computing entailments suggest that (as Montague's work also indicates) it pays to take surface structure seriously!

Cons: Language is rife with indexicals such as *I*, *you*, *here*,

now, tomorrow, this, remote, foreigner, etc., whose meaning depends on who is addressing whom, and when and where this is taking place, among other things. In addition, many words such as *bank, letter, have, cool, settlement, etc.*, are ambiguous even as fixed parts of speech. This can easily lead to faulty inferences, such as, *Mary had a little lamb* \models *Mary ate a bit of meat*. Further, coreference is problematic in such approaches; e.g., *That caused an uproar* probably refers to something reported in the preceding sentence (which could be almost anything), and the participants in the uproar also need to be inferred from prior discourse or general knowledge. There are (currently) also severe logical inference gaps, such as inability to infer the second disjunct of a binary disjunction, given the negation of the first (the classical disjunctive syllogism).

Conceptual meaning representations (e.g., Schank & Abelson (1977), Jackendoff (1990), Baker *et al.* (1998)); these are canonical representations of meanings in terms of “conceptual primitives” such as CAUSE, GO, PTRANS (physically transport) or MOVE, based on systematic study of the apparent commonalities among lexical meanings; typical sentential representations involve assigning roles such as *causal agent, theme, patient, or recipient* to arguments of verbs, and decomposing verb meanings to reflect underlying primitives; for instance, *Bill went into the house* might be represented as something like

((BILL PTRANS BILL) TO-LOC (INSIDE-OF HOUSE))

or perhaps

[EVENT GO ([AGENT BILL], [PATH TO ([PLACE IN ([THING HOUSE])])])],

depending on the particular choices of primitives, and conventions in marking roles and entity types. Importantly, some conceptual-semantics theories also posit larger-scale conceptual structures, variously called *scripts, schemas, or frames*, intended to characterize stereotyped configurations of actions and events, such as commercial transactions in which a seller transfers goods to a buyer in exchange for money, or a restaurant patron goes through the usual steps involved in restaurant dining (with participation by restaurant personnel, and use of a table, chair, etc.)

Pros: Assuming that ordinary language can be systematically mapped to primitive-based representations, paraphrastic variety may be reduced, simplifying inference. As well, to the extent that inference rules are tied to a small set of primitives, the number of rules needed for broad-coverage inference may be greatly reduced. Most importantly, perhaps, the larger-scale schemas can provide a means for rapidly recognizing what is going on in a story or other discourse, helping both with sentential interpretation and with inference of unmentioned events and participating entities.

Cons: The decompositional approach tends to lose subtle aspects of meaning; for example, Schank’s representation of *walking* as PTRANSing oneself by MOVEing one’s legs could equally well apply to running, hopping, skating, etc. As well, decomposition, coupled with schema/script instantiation, can lead to an explosion of primitive concep-

tualizations for simple sentences such as *John dined at a restaurant*, which would then greatly encumber answering a question such as *Did John dine at a restaurant?*, especially if the primitive conceptualizations from the original input were incorporated into a full-scale memory harboring millions of similar conceptualizations. Another issue is that formal interpretability is typically treated as irrelevant; but if we cannot say, for example, what in the world AGENT or PATH TO are intended to correspond to, then we cannot be sure that the larger semantic expressions we are constructing are meaningful, and that ultimately the conceptual forms assigned to sentences are true or false in the world whenever the NL sentences they were derived from were true or false in the world; and if that is the case, we have no basis for judging whether extended inferencing, beyond immediate consequences of the primitive conceptualizations derived from language, will lead to sensible conclusions or nonsense.

Thematic role representations (related to preceding item). These can be thought of as less radical departures from surface form than conceptual representations, retaining role notions such as agent, theme, recipient, location, etc., in representing verbal (and sometimes other) relations detected in parsed text, but not attempting decomposition into primitives or fitting the detected relations into stereotyped schemas.

Pros: Thematic role labeling lends itself well to ML techniques, perhaps synergistically with parse disambiguation and word sense disambiguation (Palmer, Gildea, and Xue 2010) and may prove to be of significant value in machine translation (since intuitively roles involved in a given sentence should be fairly invariant across languages).

Cons: However, as a meaning representation, parsed sentences with role labels fall short in various ways, for instance making no real attempt to adequately capture the meaning of modification, quantification, or reification, and thus does not support a theory of entailment and inference in any strong sense.

FOL: predication + connectives + quantifiers + equality; this probably needs no further explanation. (Standard computational linguistics texts such as (Allen 1995) or (Jurafsky and Martin 2009) provide introductions to FOL-based NL interpretation.)

Pros: FOL is sufficiently expressive to capture the meaning of many English sentences, especially if these are concerned with objective, cut-and-dried physical matters such as employee rosters, product inventories, flight or hotel booking, purchases, and the like, rather than “human concerns” such as beliefs, desires, or plans, or imprecise or uncertain knowledge. FOL is well-understood, and inference machinery for it is obtainable off-the-shelf.

Cons: However, we have already noted expressive weaknesses in relation to NL. There are some ways of coercing modal talk, and even talk about propositions, into FOL, for example by expressing modal notions such as necessity or belief in terms of quantification over possible worlds; or by “functionalizing” all formulas, and using a 1-place *Holds*

or *True* predicate to assert that a functionalized formula is true. Thus we might say *Holds (loves(Romeo,Juliet))* instead of *Loves(Romeo,Juliet)*, where *loves*, in contrast with *Loves*, is a function that forms a propositional entity. However, such approaches become awkward when extended to full linguistic coverage: Modals don't cover all forms of intensionality, and the *Holds* device runs into trouble for embedded quantifiers, as in the sentence

Kim believes that every galaxy harbors life.

The problem is that quantifiers can't be functionalized, because functions don't bind variables. McCarthy (1990) had a tentative proposal concerning this problem, but it did not address generalized quantifiers or event reference, and his overall scheme involved considerable complications (concept functions, denotation functions, substitution functions, and other devices). The overarching question about such approaches is whether the representational complications, which render putative interpretations of English sentences quite hard to comprehend, are really warranted just to be able to deploy FOL technology (which is probably not optimized for functionalized FOL in any case). As argued earlier, nothing will be gained in terms of tractability. Ultimately, the tool should be adapted to the task, rather than the other way around.

DRT (Discourse Representation Theory): nested discourse structures with free variables, interpreted dynamically; this approach was developed by Kamp (1981) and Heim (1982) largely to address two issues concerning anaphoric pronouns: what principles govern binding of such pronouns (i.e., what prior noun phrases they can and cannot refer to); and how "donkey pronouns" could be interpreted in an intuitively satisfactory way:

If John owns a donkey, he will ride it in the parade.

If we interpret "a donkey" as an existentially quantified term (as we surely would if *John owns a donkey* stood alone) then, if we retain narrow scope for the quantifier, we won't be able to bind the pronoun *it* to the donkey-variable, as it lies outside the scope of the quantifier. If instead we scope the existential quantifier over the entire sentence, we obtain an unwanted reading, viz., there is some donkey such that if John owns it, he will ride it. But this will be true as long as there is some donkey that he doesn't own!

DRT represents indefinites as free variables, but uses a state-change semantics (where states are assignments of values to variables) that is capable of carrying earlier variable bindings forward to later occurrences of those variables. The details need not concern us, because DR structures derived from sentences can be converted to FOL. This is the method used by Johan Bos' Boxer system for semantic parsing (Bos 2008).

Pros: Before conversion to FOL, DR structures allow for systematic determination of "accessible" referents for anaphoric pronouns. After conversion to FOL, standard FOL inference technology can be applied.

Cons: The same expressivity issues as for FOL apply.

Semantic networks: These are graphical representations of

predicate-argument structure or operator-operand structure, avoiding duplication of identical terms and other subexpressions. Every distinct term or subexpression corresponds to a distinct node, and there are labeled pointers from each such node to the nodes corresponding to its parts. (See the articles in Sowa, 1991.) For example, in the representation of *Nemesis and Liriope both believe that Narcissus loves himself*, the node for the *love*-proposition would have both a subject pointer and an object pointer to the node for *Narcissus*, and the two *believe*-propositions would both have object pointers to the *love*-proposition. There are various versions of semantic networks, many of which slightly exceed FOL in expressivity.

Pros: The main advantage of semantic networks is that graphical representations that don't duplicate subexpressions may suggest both ways of indexing knowledge (e.g., making propositions about an entity accessible from the node representing it) and ways of making inferences one might not otherwise think of (such as "marker-passing" schemes that exploit the transitivity of "isa" relations and other entailment-like relations).

Cons: As more and more propositions are added to a semantic network diagram, it tends to become an unreadable tangle of crossing and swerving lines. Of course, that is no impediment in terms of internal representations in a computer, and in fact many nominally *non*-network representations adopt a "structure-sharing" strategy that names all subexpressions and never duplicates the same structure. So the real limitations of semantic networks just lie in whatever expressive limitations they may have, or (less often) in lacking any semblance of a model theory. Thus they might lack disjunction, or generalized quantification, etc.; or they might consist of triples such as (*Man own Dog*) or (*Dog (best friend-of) Man*), where it is up to the reader to decide whether these are intended as claims about particular individuals or generic claims.

Description logics (DLs), such as OWL-DL: These generally consist of a terminological language + assertion language, where both are really aimed at supporting knowledge engineering in restricted domains, rather than NL interpretation (e.g., Cimiano *et al.* (2014)). As a semantic representation for linguistic content, the OWL-DL assertion language is entirely inadequate, allowing little more than ground predications; and even the terminological language is weak, if broad coverage is sought.

Hobbs' LF (Hobbs 2006): This is a "flat" representation in which all words are treated as predicates, typically containing ordinary individuals as well as "eventuality" arguments (events or propositions) they are considered to refer to. *Pros:* Hobbs allows everything into the domain of individuals that can be referred to by a noun phrase; that gives it breadth of coverage and provides referents for referring expressions. In principle it also allows application of FOL inference machinery. *Cons:* The reduction of all functional types—including quantifiers, connectives, and modifiers—to predicates, the approach to reification, and the amalgamation of events and propositions, seem deeply problematic:

- Conflation of propositions with events; this is illicit, witness, *The neighbors were woken up by John('s) firing a gun last night*, vs. #*The neighbors were woken up by the proposition that John fired a gun last night*.
- Quantification is represented using “typical elements” of sets that are not members of those sets but have all the properties *shared* by the set elements, and no others; this can be shown to lead to contradiction, for instance by considering the set $S = \{0,1\}$, whose members share the property of being in S ; thus the typical element is also in S , yet by definition, is not in S ;
Also this approach to quantification doesn't allow statements of type *All members x of set S are R -related to all other members y of set S* , as this will relate the typical element to itself.
- Failure to distinguish sentential and predicate modifiers; e.g., in *Skeptically, John scrutinized the proof of Fermat's last theorem*, vs. *Undoubtedly, John scrutinized the proof of Fermat's last theorem*.

As well, as a SR, Hobbs' LF does not do well on the criterion of intuitive understandability, as a result of neglecting all type distinctions and deriving “flat” rather than nested LFs.

Abstract Meaning Representation (AMR) (Banarescu et al. 2013): This has some similarity to Hobbs, blended with ideas from conceptual and frame representations; for example, *the destruction of the room by the boy* receives the same SR as *the boy destroyed the room*:
(*d / destroy-1 :arg0 (b / boy) :arg1 (r / room)*).

Similarly *to fear* and *to be afraid of* are reduced to the same *fear-01* semantic frame. The free variables are perhaps implicit wide-scope existentials, though this is left open. Notably, there is no reliance on a syntactic parse; the idea is that such LFs should be learned from an SR-annotated corpus of otherwise unanalyzed sentences.

Pros: To the extent that the reduction of deverbal nouns and adjectives to their verbal basis can actually be learned, a kind of automated normalization might be achieved; an initial result is reported in (Braune, Bauer, and Knight 2014); application to semantics-based MT has also been initiated (Jones et al. 2012).

Cons: As semantics, this is very rough; eg., *The girl adjusted the machine* and *The girl made adjustments to the machine* are assigned the same representation. So how would *The girl made several major adjustments and many minor adjustments to the machine* be represented? The AMR authors concede that quantification is not adequately handled, and hypothetical or imagined events are not distinguished; in other words, intensionality is ignored for now; there is no formal semantics, and so it is left to informal intuition exactly what is meant by plugging entities into slots such as *:time*, *:domain*, *:mod*, or *:manner*. It seems unlikely that this SR will lend itself to effective inferencing, if it is feasible at all.

Intensional logic (Montagovian, or Montague-inspired) (Dowty 1979); Montague treats language as an indexical

intensional logic, with a strictly compositional semantics, i.e., the semantic denotation of any phrase is obtained by function composition of the denotations of its immediate (top-level) constituents; there are infinitely many types; NP meanings are uniformly construed as 2nd-order monadic predicates, applicable to 1st-order monadic predicates;

Pros: Montague develops a beautifully uniform correspondence between syntactic types and semantic types; intuitions about entailments of intensional locutions are accurately captured;

Cons: The base semantics is counterintuitive, e.g., names as referring to properties of properties; Montague introduced separate names for the entities themselves. The handling of quantifier scope ambiguity is awkward, based on generating alternative parses that front different NPs; this approach doesn't extend to coordinator scope ambiguity (as in *Every boy loves Peggy or Sue*). Inference is feasible but not particularly elegant, involving much adding and dropping of intension (\wedge) and extension (\vee) operators¹ and λ -manipulation.

Montague-like compositional approaches, minus intensionality (typically retaining use of λ -calculus) (McAllester and Givan 1992; Artzi and Zettlemoyer 2013; Kwiatkowski et al. 2013; Howard, Tellex, and Roy 2013); *Pros*: Elegant compositional semantics, as long as we do not venture into intensional locutions; ease of integration with feature-based, trainable disambiguation, learnability of the transduction.

Cons: expressive limitations, no events or times, etc.

Situation semantics (SS) (Barwise and Perry 1983); SS semantics was intended as a “realistic” semantics that represents the information contents of situations in terms of concrete relations – tuples of properties/relations and their participants, with positive or negative polarity; possible worlds (central to Montague semantics) are avoided in favor of situation types that are themselves part of the ontology.

Pros: Events or situations can be described by arbitrarily complex sentences; the semantics broadens the notion of context to allow (in effect) for the mental state of a perceiver in perception statements. Cooper (2005) imports some key ideas from SS into a DRT-like framework (with feature-structured “records” instead of DRSs), and shows examples of compositional LF computation for a few sentences, including “donkey sentences”.

Cons: The metaphysics is very abstruse and complex (seemingly as the result of avoiding *possibilia*), the status of attitudes and other forms of intensionality remains somewhat unsettled, and the relationship of SS to the semantic representation of linguistic content has not been very extensively worked out. It is unclear how inference is to be performed in an SS-based SR, given the dynamic, context-dependent semantics in SS.

¹For those acquainted with Montague Grammar: This complication can be avoided by making possible worlds the last, rather than the first, argument of interpretation functions; this is the strategy in *episodic logic* (also considered here).

Episodic logic (EL): This is a Montague-inspired, language-like SR (and general knowledge representation) (Schubert and Hwang 2000; Schubert 2000; Hwang and Schubert 1994; Schubert 2014); it is first-order, intensional, and strictly compositional, apart from initially unscoped quantifiers, logical connectives and tense; this is reflected in very simple LF formation rules.

Pros: EL accounts very simply for quantifier and connective ambiguity. Instead of using arbitrarily high-order types, EL uses type-shifting operators that “lower” predicates and sentences to individuals; in this way it handles all forms of intensionality, as well as generalized quantifiers, etc. It treats events/situation (episodes) realistically, and they can therefore be characterized in arbitrarily many, specific or abstract ways. Integration with specialists is straightforward. Two well-developed inference engines exist, the latest being EPILOG 2 (Morbini and Schubert 2009; 2011); the former publication shows EPILOG 2 to be competitive in the FOL theorem-proving domain, despite its handling of a much enriched representation. Inference in EPILOG is fundamentally similar to NLog (but more general, see Schubert (2014b)), which is made possible by the close correspondence between EL’s expressive devices and those of NL.

Cons: Inference remains brittle, and probabilistic inference is not implemented in a theoretically and practically satisfactory way.

Conclusions

In this brief survey of broad-coverage SRs, it has not been possible to evaluate each of the representational styles with respect to each desideratum – this would have required some 120 capsule commentaries. However, we can make the following general observations (where phrases referring to the initial desiderata are italicized).

NL-like expressivity is attained in few SRs that support *inferences* (beyond similarity judgements or type subsumption). Structured English, Montague-inspired approaches, EL, and Hobbs’ LF come closest, though the first of these must ultimately address indexicality issues and inference limitations. Adequate *referent availability*, including reference to events, facts, etc., seems to be limited to a few “ontologically promiscuous” representation. *Formal interpretability* is an issue for structured English, conceptual SRs, thematic role representations (if taken beyond simple predications), some types of semantic networks, and AMR. Weakness in this area also tends to correlate with lack of inference mechanisms (though not entirely in the case of structured English or certain conceptual representations). Judging whether putative semantic representations of phrases and sentences *accord with intuitions* about their meaning and entailments is difficult or uncertain for many SRs; structured English and representations whose surface form is close to NL (e.g., Montagovian, or EL) fare best in this respect.

Ease of mapping from NL to the SR is also simplest and most transparent for NL-like SRs – often semantic rules are just a matter of applying the semantic functor corresponding to the head of a phrase to the semantic interpretations of the complement. *Trainability* so far is easiest for the

least ambitious SRs, in terms of the semantic aspects of language being captured, or the domain addressed. Simple thematic role SRs, strictly limited non-intensional Montague-like SRs, and AMRs fare best so far. However, machine learning of the mapping from NL to SR is hardest, requiring myriad training examples, when the structural gap between the NL source and SR target is greatest. Given that even structured English readily lends itself to inference, it seems very worthwhile to attempt learning of semantic parsers for comprehensive, NL-like, but more formal SRs.

In short, broad, language-like, inference-enabling SRs in the spirit of Montague are feasible, while avoiding Montague’s unbounded type theory and ubiquitous intension/extension operators. Such a representation, exemplified by EL, also reflects Hobbs’ advocacy of ontological promiscuity, without trying to avoid natural type distinctions, such as the distinction between quantifiers and predicates – which surely is needed in a general world knowledge base supporting language understanding in any case.

References

- Allen, J. F. 1995. *Natural Language Understanding*. Redwood City, CA: Benjamin/Cummings.
- Angeli, G., and Manning, C. 2014. NaturalLI: Natural Logic inference for common sense reasoning. In *Empirical Methods in Natural Language Processing (EMNLP)*.
- Artzi, Y., and Zettlemoyer, L. 2013. Weakly supervised learning of semantic parsers for mapping instructions to actions. *Trans. of the Assoc. for Computational Linguistics (TACL)*.
- Artzi, Y.; Kwiatkowski, T.; and Berant, J., eds. 2014. *ACL 2014 Workshop on Semantic Parsing (SP’14)*.
- Baker, C.; Fillmore, C.; and Lowe, J. 1998. The Berkeley framenet project. In *Ann. Meet. of the Assoc. for Computational Linguistics and the Int. Conf. on Computational Linguistics (ACL-COLING ’98)*, 86–90.
- Banarescu, L.; Bonial, C.; Cai, S.; Georgescu, M.; Griffitt, K.; Hermjakob, U.; Knight, K.; Koehn, P.; Palmer, M.; and Schneider, N. 2013. Abstract meaning representation for sembanking.
- Barwise, J., and Perry, J. 1983. *Situations and Attitudes*. Cambridge, MA: MIT Press.
- Berant, J.; Dagan, I.; and Goldberger, J. 2011. Global learning of typed entailment rules. In *Ann. Meet. of the Assoc. for Computational Linguistics: Human Language Technologies, Vol. 1 (HLT’11)*.
- Bos, J. 2008. Wide-coverage semantic analysis with Boxer. In Bos, J., and Delmonte, R., eds., *Semantics in Text Processing (STEP 2008)*, 277–286. College Publications.
- Braune, F.; Bauer, D.; and Knight, K. 2014. Mapping between English strings and reentrant semantic graphs. In *Int. Conf. on Language Resources and Evaluation (LREC)*.
- Cimiano, P.; Unger, C.; and McCrae, J. 2014. *Ontology-Based Interpretation of Natural Language*. Morgan & Claypool Publishers.

- Clark, S. 2013. Vector space models of lexical meaning. In Lappin, S., and Fox, C., eds., *Handbook of Contemporary Semantics*, 2nd ed. Malden, MA: Blackwell.
- Clark, P. 2015. Elementary school science and math tests as a driver for AI: Take the Aristo Challenge! to appear.
- Clarke, D. 2012. A context-theoretic framework for compositionality in distributional semantics. *Computational Linguistics* 38(1):41–71.
- Cooper, R. 2005. Records and record types in semantic theory. *Journal of Logic and Computation* 15(2):99–112.
- Dagan, I.; Bar-Haim, R.; Szpektor, I.; Grental, I.; and Shnarch, E. 2008. Natural language as the basis for meaning representation and inference. In Gelbukh, A., ed., *Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science 4919*. Berlin: Springer. 151–170.
- Dagan, I.; Glicksman, O.; and Magnini, B. 2006. The PAS-CAL recognising textual entailment challenge. In nonero-Candela et al., J. Q., ed., *MLCW 2005, LNAI 3944*. Heidelberg: Springer.
- Dowty, D. R. 1979. *Word Meaning and Montague Grammar*. Boston, MA: D. Reidel Publ. Co.
- Eliasson, K. 2007. Case-based techniques used for dialogue understanding and planning in a human-robot dialogue system. In *Proc. of the 20th Int. Joint Conf. on Artificial Intelligence (IJCAI'07)*, 1600–1605.
- Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. Dissertation, University of Massachusetts, Amherst.
- Hobbs, J. R. 2006. Discourse and inference. In progress at <http://www.isi.edu/~hobbs/disinf-tc.html>.
- Howard, T.; Tellex, S.; and Roy, N. 2013. A natural language planner interface for mobile manipulators. In *30th Int. Conf. on Machine Learning, JMLR: W&CP volume 28*.
- Hwang, C., and Schubert, L. 1994. Interpreting tense, aspect, and time adverbials: a compositional, unified approach. In Gabbay, D., and Ohlbach, H., eds., *Proc. of the 1st Int. Conf. on Temporal Logic*. Bonn, Germany: Springer. 238–264.
- Jackendoff, R. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.
- Jones, B.; Andreas, J.; Bauer, D.; Hermann, K.-M.; and Knight, K. 2012. Semantics-based machine translation with hyperedge replacement grammars. In *Proc. of the Int. Conf. on Computational Linguistics (COLING'12)*.
- Jurafsky, D., and Martin, J. H. 2009. *Speech and Language Processing (2nd Edition)*. Upper Saddle River, NJ: Prentice-Hall.
- Kamp, H. 1981. A theory of truth and semantic representation. In Groenendijk, J.; Janssen, T.; and Stokhof, M., eds., *Formal Methods in the Study of Language*. Mathematical Centre-tracts, U. Amsterdam. 277–320.
- Kwiatkowski, T.; Choi, E.; Artzi, Y.; and Zettlemoyer, L. 2013. Scaling semantic parsers with on-the-fly ontology matching. In *Empirical Methods in Natural Language Processing (EMNLP)*.
- Levesque, H.; Davis, E.; and Morgenstern, L. 2012. The Winograd Schema Challenge. In *Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR'12)*.
- MacCartney, B., and Manning, C. 2009. An extended model of natural logic. In *8th Int. Conf. on Computational Semantics (IWCS-8)*, Tilburg University, Netherlands.
- McAllester, D., and Givan, R. 1992. Natural language syntax and first order inference. *Artificial Intelligence* 56:1–20.
- McCarthy, J. 1990. First order theories of individual concepts and propositions. In McCarthy, J., and Lifschitz, V., eds., *Formalizing Common Sense: Papers by John McCarthy*.
- Melamud, O.; Berant, J.; Dagan, I.; Goldberger, J.; and Szpektor, I. 2013. A two level model for context sensitive inference rules. In *Ann. Meet. of the Assoc. for Computational Linguistics (ACL'13)*.
- Morbini, F., and Schubert, L. K. 2009. Evaluation of EPILOG: a reasoner for Episodic Logic. In *Commonsense 09*.
- Morbini, F., and Schubert, L. 2011. Metareasoning as an integral part of commonsense and autocognitive reasoning. In Cox, M., and Raja, A., eds., *Metareasoning: Thinking about Thinking, Ch. 17*. MIT Press. 267–282.
- Palmer, M.; Gildea, D.; and Xue, N. 2010. *Semantic Role Labeling*. Morgan & Claypool.
- Roemmele, M.; Bejan, C.; and Gordon, A. 2011. Choice of plausible alternatives: An evaluation of commonsense causal reasoning. In *Int. Symp. on Logical Formalizations of Commonsense Reasoning*.
- Schank, R., and Abelson, R. 1977. *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Lawrence Erlbaum.
- Schubert, L. K., and Hwang, C. H. 2000. Episodic Logic Meets Little Red Riding Hood: A comprehensive, natural representation for language understanding. In Iwanska, L., and Shapiro, S., eds., *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language*, 111174. Menlo Park, CA, and Cambridge, MA: MIT/AAAI Press.
- Schubert, L. K. 2000. The situations we talk about. In Minker, J., ed., *Logic-Based Artificial Intelligence*. Kluwer, Dordrecht. 407–439.
- Schubert, L. 2014. From Treebank parses to Episodic Logic and commonsense inference. In *Workshop on Semantic Parsing (SP'14)*.
- Schubert, L. 2014b. Nlog-like inference and commonsense reasoning. In Zaenen, A.; de Paiva, V.; and Condoravdi, C., eds., *Perspectives on Semantic Representations for Textual Inference, special issue of Linguistics Issues in Language Technology – LiLT 9(9)*. Stanford, CA: CSLI Publications.
- Sowa, J., ed. 1991. *Principles of Semantic Networks: Explorations in the Representation of Knowledge*. San Mateo, CA: Morgan Kaufmann.
- Widdows, D. 2004. *Geometry and Meaning*. Stanford, CA: CSLI Publications.